



# Sistematización y métodos estadísticos activ 2

- Rojas Cortez Nayely

**DOCENTE**

Castro Lopez, Segundo Vicente

# GRUPO 5 - CIRROSIS

- Rojas Cortez Nayely
- Ticse Pecho Katherine
- Manay Ventura Alexander
- Pachas Rojas Lizbeth
- Lozano Laura,Marina
- Caballero Martinez Arianna

# Instalación de paquetes

```
{r}
#PASO 1 INSTALAR SOLO SI NO ESTA INSTALADOS - SI SALE ERROR ES PORQUE YA ESTA
INSTALADO

install.packages("tidyverse")
install.packages("rio")
install.packages("here")
install.packages("janitor")
```

```
{r}
#PASO 2 ACTIVAR PAQUETES DE FORMA OBLIGATORIA

library(tidyverse)
library(rio)
library(here)
library(janitor)

— Attaching core tidyverse packages
tidyverse 2.0.0

—
✓ dplyr      1.1.4      ✓ readr      2.1.5
✓forcats     1.0.0      ✓ stringr    1.5.1
✓ ggplot2    3.5.2      ✓ tibble     3.3.0
✓ lubridate   1.9.4      ✓ tidyr     1.3.1
✓ purrr     1.1.0

— Conflicts
— tidyverse_conflicts() —
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
i Use the conflicted package to force all conflicts to become errors

Some optional R packages were not installed and therefore some file formats are
not supported. Check file support with show_unsupported_formats()
here() starts at C:/Users/Forest/Documents/METODOS ESTADISTICOS/estadistica_upsjb

Adjuntando el paquete: 'janitor'

The following objects are masked from 'package:stats':

  chisq.test, fisher.test
```

Es importante activar los paquetes para un correcto uso en Rstudio

```
{r}  
cirrosis= import(here("data", "cirrosis.csv"))
```

```
Registered S3 method overwritten by 'data.table':  
  method           from  
print.data.table
```

Se importa la data base a modificar, a nuestro grupo le tocó cirrosis

```
{r}
str(cirrosis)

'data.frame': 418 obs. of 20 variables:
 $ ID           : int 1 2 3 4 5 6 7 8 9 10 ...
 $ Dias_Seguimiento : int 400 4500 1012 1925 1504 2503 1832 2466 2400 51 ...
 $ Estado        : chr "Fallecido" "Censurado" "Fallecido" "Fallecido" ...
 $ Medicamento   : chr "D_penicilamina" "D_penicilamina" "D_penicilamina" ...
 "D_penicilamina" ...
 $ Edad          : int 21464 20617 25594 19994 13918 24201 20284 19379 15526
25772 ...
 $ Sexo          : chr "Mujer" "Mujer" "Hombre" "Mujer" ...
 $ Ascitis       : chr "Sí" "No" "No" "No" ...
 $ Hepatomegalia: chr "Sí" "Sí" "No" "Sí" ...
 $ Aracnoides    : chr "Sí" "Sí" "No" "Sí" ...
 $ Edema         : chr "Severo" "Ausente" "Leve" "Leve" ...
 $ Bilirrubina   : num 14.5 1.1 1.4 1.8 3.4 0.8 1 0.3 3.2 12.6 ...
 $ Colesterol    : int 261 302 176 244 279 248 322 280 562 200 ...
 $ Albumina      : num 2.6 4.14 3.48 2.54 3.53 3.98 4.09 4 3.08 2.74 ...
 $ Cobre         : int 156 54 210 64 143 50 52 52 79 140 ...
 $ Fosfatasa_Alcalina: num 1718 7395 516 6122 671 ...
 $ SGOT          : num 137.9 113.5 96.1 60.6 113.2 ...
 $ Trigliceridos: int 172 88 55 92 72 63 213 189 88 143 ...
 $ Plaquetas     : int 190 221 151 183 136 NA 204 373 251 302 ...
 $ Tiempo_Protrombina: num 12.2 10.6 12 10.3 10.9 11 9.7 11 11 11.5 ...
 $ Etapa         : chr "Etapa 4" "Etapa 3" "Etapa 4" "Etapa 4" ...
```

Con str se  
puede ver  
un  
resumen de  
los datos de  
cada  
variable

```
{r}
cirrosis_1 = cirrosis |>
  clean_names() |>
  mutate_if(is.character, list(~ na_if(., ""))) |>
  remove_empty(which = c("rows", "cols"))
```

- Nombres de columnas: Convertidos a minúsculas con guiones bajos (clean\_names())
- Celdas vacías: Reemplazadas por NA en todas las columnas de texto
- Filas/columnas vacías: Eliminadas completamente del dataset

```
{r}
library(dplyr)

# Código corregido con nombres en minúsculas
cirrosis_2 <- cirrosis_1 %>%
  # 1. Convertir edad de días a años y redondear
  mutate(edad = round(edad / 365.25)) %>%
  
  # 2. Estandarizar "Sí"/"No" en columnas categóricas binarias
  mutate(across(
    c(ascitis, hepatomegalia, aracnoides),
    ~ case_when(
      tolower(.x) %in% c("sí", "si", "yes", "s", "y") ~ "SI",
      tolower(.x) %in% c("no", "n") ~ "NO",
      TRUE ~ as.character(.x)
    )
  )) %>%
  
  # 3. Estandarizar valores de edema
  mutate(edema = case_when(
    tolower(edema) %in% c("leve", "ligero", "moderado") ~ "Leve",
    tolower(edema) == "severo" ~ "Severo",
    tolower(edema) == "ausente" ~ "No",
    TRUE ~ as.character(edema)
  ))
  
# Verificar resultados
glimpse(cirrosis_2)
summary(cirrosis_2)
```

- EDAD: Convertida de días a años (valores enteros)
- COLUMNAS BINARIAS: Estandarizadas a "Sí"/"No" (ascitis, hepatomegalia, aracnoides)
- EDEMA: Categorizado en "Leve"/"Severo"/"No"
- DATOS FALTANTES: Valores como "ausente" convertidos a "No"
- CONSISTENCIA: Minúsculas/mayúsculas uniformizadas en respuestas

```
{r}  
str(cirrosis_2)
```

  

```
'data.frame': 418 obs. of 20 variables:  
 $ id : int 1 2 3 4 5 6 7 8 9 10 ...  
 $ dias_seguimiento : int 400 4500 1012 1925 1504 2503 1832 2466 2400 51 ...  
 $ estado : chr "Fallecido" "Censurado" "Fallecido" "Fallecido" ...  
 $ medicamento : chr "D_penicilamina" "D_penicilamina" "D_penicilamina"  
 "D_penicilamina" ...  
 $ edad : num 59 56 70 55 38 66 56 53 43 71 ...  
 $ sexo : chr "Mujer" "Mujer" "Hombre" "Mujer" ...  
 $ ascitis : chr "SI" "NO" "NO" "NO" ...  
 $ hepatomegalia : chr "SI" "SI" "NO" "SI" ...  
 $ aracnoides : chr "SI" "SI" "NO" "SI" ...  
 $ edema : chr "Severo" "No" "Leve" "Leve" ...  
 $ bilirrubina : num 14.5 1.1 1.4 1.8 3.4 0.8 1 0.3 3.2 12.6 ...  
 $ colesterol : int 261 302 176 244 279 248 322 280 562 200 ...  
 $ albumina : num 2.6 4.14 3.48 2.54 3.53 3.98 4.09 4 3.08 2.74 ...  
 $ cobre : int 156 54 210 64 143 50 52 52 79 140 ...  
 $ fosfatasa_alcalina: num 1718 7395 516 6122 671 ...  
 $ sgot : num 137.9 113.5 96.1 60.6 113.2 ...  
 $ trigliceridos : int 172 88 55 92 72 63 213 189 88 143 ...  
 $ plaquetas : int 190 221 151 183 136 NA 204 373 251 302 ...  
 $ tiempo_protrombina: num 12.2 10.6 12 10.3 10.9 11 9.7 11 11 11.5 ...  
 $ etapa : chr "Etapa 4" "Etapa 3" "Etapa 4" "Etapa 4" ...
```

Usando str  
con la data  
base ya limpia  
para poder  
observar los  
cambios

```
{r}
# Exportar como CSV
write.csv(cirrosis_2, "cirrosis_2.csv", row.names = FALSE, fileEncoding = "UTF-8")

# Mensaje de confirmación
cat("Base de datos exportada como: cirrosis_2.csv")
```

Base de datos exportada como: cirrosis\_2.csv

Se exporta la data base en  
la carpeta scripts

Muchas  
Gracias

